文章编号:1005-3085(2011)02-0211-09

纵向数据下部分线性EV模型的变量选择*

杨宜平[1]， 薛留根[2]， 程维虎[2]

(1- 重庆工商大学数学与统计学院，重庆 400067； 2- 北京工业大学应用数理学院，北京 100124)

摘 要: 本文考虑了纵向数据下部分线性EV模型的变量选择问题，采用SCAD惩罚方法提出了一个变量选择过程. 通过选择适当的惩罚参数，证明了该变量选择过程可以相合地识别出真实模型，并且所得的正则估计具有Orcale性质. 最后模拟研究了所提出方法的有限样本性质.

关键词: 部分线性EV模型；估计理论；数据分析；SCAD惩罚；Orcale性质

分类号: **AMS**(2000) 62G05; 62G20 **中图分类号**: O212.7 **文献标识码**: A

1 引言

考虑来自n个个体的数据，其第i个个体具有m_i次观测$i=1,2,\cdots,n$，总的观测数为$N=\sum\limits_{i=1}^{n}m_i$. 设$Y_{ij}$和$(X_{ij},T_{ij})$分别是第$i$个个体的第$j$次观测$j=1,2,\cdots,m_i$的响应变量和协变量，其中$X_{ij}$是$p\times 1$向量而$T_{ij}$是数量或时间. 响应变量和协变量的依赖关系由下式给出

$$Y_{ij}=X_{ij}^T\beta+g(T_{ij})+\varepsilon_{ij}, \tag{1}$$

其中β是$p\times 1$未知回归系数向量，$g(t)$是未知的基准函数，ε_{ij}是随机误差，且满足

$$\mathrm{E}\big(\varepsilon_{ij}\,|\,X_{ij},T_{ij}\big)=0 \quad 且 \quad \sigma_\varepsilon^2(t)=\mathrm{E}\big(\varepsilon_{ij}^2\,|\,T_{ij}=t\big).$$

不失一般性，假定T_{ij}在闭区间$[0,1]$上取值，进一步假定来自不同个体的观测独立.

本文感兴趣的是模型(1)中协变量X_{ij}不能直接观测到的情形，即协变量带有测量误差

$$W_{ij}=X_{ij}+U_{ij}, \quad i=1,2,\cdots,n, \quad j=1,2,\cdots,m_i, \tag{2}$$

其中，测量误差U_{ij}为i.i.d.随机变量，与(Y_{ij},T_{ij},X_{ij})独立，均值为零，协方差阵为Σ_{uu}. 式(1)和式(2)被称为部分线性测量误差模型或部分线性EV(errors-in-variables)模型. 本文考虑Σ_{uu}已知的情形，当Σ_{uu}未知时，可以采用Carroll等[1]提出的方法获得Σ_{uu}相合的，无偏的矩估计，本文的结论仍成立.

实际上，由于各种原因，测量误差是普遍存在的，因此，研究模型(1)和(2)具有实用价值. 不少文献已经讨论了部分线性EV模型：崔恒建[2]考虑了有重复观测的部分线性EV模型，给出了诸多估计及估计的性质；Liang等[3]考虑了部分线性EV模型，在测量误差协方差已知的情形下，获得了参数和非参数估计及渐近性质；Zhu和Cui[4]考虑了参数部分和非参数部分均含有误差的情形，通过利用矩和反卷积方法给出了未知参数的估计，并给出了估计的相

收稿日期: 2009-04-14. **作者简介**: 杨宜平(1981年12月生)，女，博士. 研究方向：非参数统计.

***基金项目**: 国家自然科学基金(10871013)；高等学校博士学科点专项科研基金(20070005003)；北京市自然科学基金(1102008)；北京市属高等学校人才强教计划资助项目；重庆工商大学科研启动项目(20105609).

合性和渐近正态性. 关于纵向数据下部分线性模型的讨论，可以参见Xue和Zhu[5]，田萍和薛留根[6]，Fan和Li[7]等的工作.

本文考虑纵向数据下部分线性EV模型同时进行变量选择和估计未知参数的问题. 目前，半参数模型的变量选择问题受到了统计学者的重视. Fan和Li[7]研究了纵向数据下半参数模型的变量选择，Liang和Li[8]考虑了部分线性EV模型的变量选择. 本文的目的是采用Fan和Li[9]提出的SCAD惩罚方法同时进行变量选择和估计未知参数. SCAD方法具有一些优点：首先，SCAD惩罚方法同时进行变量选择和估计参数，得到的估计具有Orcale性质；其次，Fan和Li[9]也给出了一个局部二次逼近的算法. 基于这些优点，SCAD惩罚已经成为一个非常流行的变量选择方法，一些相关的工作可以参见文献[7-11]. 为了变量选择，我们将β分解为$\beta=(\beta^{(1)T},\beta^{(2)T})^T$，不失一般性，假定$\beta^{(1)}$为$\beta$的前$s$个分量，为非零系数，$\beta^{(2)}$为$\beta$剩下的$p-s$个分量，为零系数.

2 方法与主要结果

下文假定m_i是有界的. 该假定意味着总的样本容量N与个体数n是同阶的量. 进一步假设$T_{ij}\,(i=1,2,\cdots,n,\ j=1,2,\cdots,m_i)$是独立同分布(i.i.d.)的，其共同的密度$f$是Lebesgue可测的. 虽然时间设计点列的i.i.d.假定理论上仅适合随机设计情形，但本文的结果和证明可以做适当修改后同样适合固定设计情形.

2.1 惩罚估计函数

在(1)式两边求给定T_{ij}下的条件期望，可得

$$Y_{ij}-\mathrm{E}\big(Y_{ij}\,|\,T_{ij}\big)=\big[X_{ij}-\mathrm{E}(X_{ij}\,|\,T_{ij})\big]^T\beta+\varepsilon_{ij}. \tag{3}$$

由(3)式，可以构造一个关于β的估计函数

$$U(\beta)=\sum_{i=1}^{n}\sum_{j=1}^{m_i}\omega(T_{ij})\big\{\check{W}_{ij}(\check{Y}_{ij}-\check{W}_{ij}^T\beta)+\Sigma_{uu}\beta\big\},$$

其中$\check{W}_{ij}=W_{ij}-\mu_1(T_{ij})$, $\check{Y}_{ij}=Y_{ij}-\mu_2(T_{ij})$且$\omega(\cdot)$是一个权函数. 这里

$$\mu_1(\cdot)=\mathrm{E}\big(W_{ij}\,|\,T_{ij}=\cdot\big),\quad \mu_2(\cdot)=\mathrm{E}\big(Y_{ij}\,|\,T_{ij}=\cdot\big).$$

定义β的惩罚估计函数为

$$U^p(\beta)=U(\beta)-nb_\lambda(\beta), \tag{4}$$

其中$b_\lambda(\beta)=q_\lambda(|\beta|)\mathrm{sgn}(\beta)$，$\lambda$是惩罚参数，$q_\lambda(\cdot)$是SCAD惩罚函数[9]，定义为

$$q_\lambda(\,|\theta|\,)=\lambda\Big\{I\big(|\theta|\le\lambda\big)+\frac{(a\lambda-|\theta|)_+}{(a-1)\lambda}I\big(\,|\theta|>\lambda\big)\Big\},\quad a>2.$$

式(4)中包含两个未知函数$\mu_1(\cdot)$和$\mu_2(\cdot)$，不能直接应用于β的统计推断. 解决这个问题的一个自然想法是在$U^p(\beta)$中用两个估计量分别代替它们. 对于固定的点$t\in[0,1]$，利用核估计法分别定义$\mu_1(t)$和$\mu_2(t)$的估计量

$$\hat{\mu}_1(t)=\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathcal{W}_{ij}(t)W_{ij},\quad \hat{\mu}_2(t)=\sum_{i=1}^{n}\sum_{j=1}^{m_i}\mathcal{W}_{ij}(t)Y_{ij},$$

其中

$$\mathcal{W}_{ij}(t) = K_h(T_{ij}-t)\Big/\sum_{k=1}^{n}\sum_{j=1}^{m_k}K_h(T_{ij}-t),$$

h 是带宽，$K_h(\cdot)=K(\cdot/h)$ 且 $K(\cdot)$ 是核函数.

在 $U^p(\beta)$ 中分别用 $\hat{\mu}_1(T_{ij})$ 和 $\hat{\mu}_2(T_{ij})$ 分别代替 $U^p(\beta)$ 中的 $\mu_1(T_{ij})$ 和 $\mu_2(T_{ij})$，可以得到 $U^p(\beta)$ 的一个估计量 $\hat{U}^p(\beta)$，即

$$\hat{U}^p(\beta) = \hat{U}(\beta) - nb_\lambda(\beta),$$

其中

$$\hat{U}(\beta) = \sum_{i=1}^{n}\sum_{j=1}^{m_i}\omega(T_{ij})\big\{\tilde{W}_{ij}\big(\tilde{Y}_{ij}-\tilde{W}_{ij}^T\beta\big)+\Sigma_{uu}\beta\big\},$$

这里

$$\tilde{W}_{ij} = W_{ij}-\hat{\mu}_1(T_{ij}),\quad \tilde{Y}_{ij} = Y_{ij}-\hat{\mu}_2(T_{ij}).$$

通过解 $\hat{U}^p(\beta)=0$，得到 β 的估计，记为 $\hat{\beta}$.

2.2　Orcale 性质

假设 β_0 是 β 的真值，令

$$a_n = \max_j\big\{q_\lambda(|\beta_{0j}|),\, \beta_{0j}\neq 0\big\},\quad b_n = \max_j\big\{q'_\lambda(|\beta_{0j}|),\, \beta_{0j}\neq 0\big\}.$$

定理 2.1　假设 $a_n\to 0$, $b_n\to 0$，第 4 节中正则条件 C1-C7 成立，那么，$\hat{U}^p(\beta)$ 存在解 $\hat{\beta}$，使得 $\hat{\beta}=\beta_0+O_p(n^{-1/2}+a_n)$.

定理 2.1 表明，如果 $a_n=O_p(n^{-1/2})$，那么存在 $\sqrt{n}$ 收敛速度的估计.

引入一些记号

$$\Sigma_\lambda = \mathrm{diag}\big\{q'_\lambda(|\beta_{01}|)\mathrm{sgn}(\beta_{01}),\cdots,q'_\lambda(|\beta_{0s}|)\mathrm{sgn}(\beta_{0s})\big\},$$

$$\Gamma = \lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\mathrm{E}\Big\{\sum_{j=1}^{m_i}\omega(T_{ij})\big[X_{ij}-\mu_1(T_{ij})\big]^{\otimes 2}\Big\},$$

$$\Xi = \lim_{n\to\infty}n^{-1}\sum_{i=1}^{n}\mathrm{E}\Big\{\sum_{j=1}^{m_i}\omega(T_{ij})\big[(X_{ij}-\mu_1(T_{ij})+U_{ij})(\varepsilon_{ij}-U_{ij}^T\beta)+\Sigma_{uu}\beta\big]\Big\}^{\otimes 2},$$

其中 $D^{\otimes 2}=DD^T$.

定理 2.2 (Oracle 性质)　假设第 4 节中的条件 C1-C7 成立，惩罚函数满足

$$\lim_{n\to+\infty}\inf\ \lim_{\theta\to 0_+}\inf\ q_\lambda(\theta)/\lambda>0,$$

$\lambda\to 0$, $n^{1/2}\lambda\to\infty$，那么，定理 2.1 中的 $\hat{\beta}=(\hat{\beta}^{(1)T},\hat{\beta}^{(2)T})^T$ 依概率 1 满足：

(i) (稀疏性 Sparsity)　$\hat{\beta}^{(2)}=0$;

(ii) (渐近正态性)

$$\sqrt{n}\,(\Gamma_{11}+\Sigma_\lambda)\big[(\hat{\beta}^{(1)}-\beta_0^{(1)})+(\Gamma_{11}+\Sigma_\lambda)^{-1}b_\lambda(\beta_0)\big]\xrightarrow{\mathcal{L}}N(0,\Xi_{11}),$$

其中 $\xrightarrow{\mathcal{L}}$ 表示依分布收敛，Ξ_{11} 和 Γ_{11} 分别由 Ξ 和 Γ 的前 s 行和 s 列组成.

定理2.2表明，所得到的估计具有稀疏性，也就是，为了减少复杂性，估计量自动把很小的系数估计成零；另一方面，当 n 充分大时，$\Sigma_\lambda = 0$ 且 $b_\lambda(\beta_0) = 0$．因此，定理2.2 (ii) 变为

$$\sqrt{n}\,\Gamma_{11}(\hat\beta^{(1)} - \beta_0^{(1)}) \xrightarrow{\mathcal{L}} N(0, \Xi_{11}).$$

注意到无论选取什么样的权函数 $\omega(\cdot)$，所提出的估计 $\hat\beta$ 具有相合性和渐近正态性．但是 $\hat\beta$ 的方差依赖权函数 $\omega(\cdot)$，最有效的权函数应选取 $1/\sigma_\varepsilon^2(t)$．$\sigma_\varepsilon^2(t)$ 往往未知，可以采用核方法估计 $\sigma_\varepsilon^2(t)$，即

$$\hat\sigma_\varepsilon^2(t) = \frac{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m_i}\left(\hat r_{ij}^2 - \hat\beta_I^T\Sigma_{uu}\hat\beta_I\right)K_{h_1}^*(t - t_{ij})}{\sum\limits_{i=1}^{n}\sum\limits_{j=1}^{m_i}K_{h_1}^*(t - t_{ij})},$$

其中 $\hat r_{ij} = y_{ij} - W_{ij}^T\hat\beta_I - \tilde g(t_{ij})$，$\hat\beta_I$ 和 $\tilde g(t)$ 是选取权函数 $\omega(t) = 1$ 时，β 和 $g(t)$ 的估计，$K_{h_1}^*(\cdot) = K^*(\cdot/h_1)$ 且 $K^*(\cdot)$ 是核函数，h_1 是带宽．

2.3 迭代算法

采用 Fan 和 Li[9] 提出的局部二次逼近方法，在非零真值 β_{0j} 的领域内，有

$$q_\lambda(|\beta_j|)\mathrm{sgn}(\beta_j) \approx \frac{q_\lambda(|\beta_{0j}|)}{|\beta_{0j}|}\beta_j.$$

在一些正则条件下，惩罚估计函数可变为

$$n^{-1/2}\hat U^p(\beta) \approx n^{-1/2}\hat U(\beta_0) - n^{1/2}\hat A(\beta - \beta_0) - n^{1/2}\Omega_\lambda(\beta_0)\beta, \tag{5}$$

其中

$$\hat A = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\omega(T_{ij})\big\{\tilde W_{ij}\tilde W_{ij}^T - \Sigma_{uu}\big\},$$

$$\Omega_\lambda(\beta) = \mathrm{diag}\big\{q_\lambda(|\beta_1|)/|\beta_1|, \cdots, q_\lambda(|\beta_p|)/|\beta_p|\big\}.$$

由 (5) 式，可以得到如下迭代算法：

步骤1　解 $\hat U(\beta) = 0$ 获得初始估计 β^0；

步骤2　$\beta^{k+1} = \beta^k + \{n\hat A + n\Omega_\lambda(\beta^k)\}^{-1}\hat U^p(\beta^k)$；

步骤3　迭代步骤2，直到 β 收敛，记 β 的最后估计为 $\hat\beta$.

SCAD 惩罚需要选择两个参数 a 和 λ，Fan 和 Li[9] 建议选择 $a = 3.7$，因此在整篇文章中，我们选取 $a = 3.7$．进一步地，类似 Fan 和 Li[9]，本文采用广义交叉核实的方法选取 λ．最小化下式

$$\mathrm{GCV}(\lambda) = \frac{\mathrm{RSS}(\lambda)/n}{\big(1 - d(\lambda)/n\big)^2}, \tag{6}$$

其中 $\mathrm{RSS}(\lambda)$ 是残差平方和，$d(\lambda) = \mathrm{tr}[\{\hat A + \Omega_\lambda(\hat\beta_\lambda)\}^{-1}\hat A^T]$ 是有效参数的个数．我们选择 $\hat\lambda = \arg\min_\lambda \mathrm{GCV}(\lambda)$.

3 模拟结果

为了实施模拟，由如下模型产生数据

$$Y_{ij} = \sin(0.5\pi T_{ij}) + X_{ij}^T\beta + \varepsilon_{ij},$$

其中 $\beta = (0.4, 1, 1.5, 2, 0, 0, 0, 0, 0, 0)^T$. 在模拟过程中，分别取 $n = 100, 200$ 和 400 个个体，每个个体具有 $m_i = 3$ 次观测，对每一种情况，实验均重复 500 次. 按照如下方式产生数据 $X_{ijr} \sim N(1, 1.5)$, $r = 1, 2, \cdots, 10$, $T_{ij} \sim U(0, 1)$. ε_{ij} 来自于 $N(0, \sigma^2)$，组内相关系数 $\rho_{\varepsilon_{ij}\varepsilon_{ik}} \equiv \rho$, $i = 1, 2, \cdots, n$, $j, k = 1, 2, 3$，Y_{ij} 由模型产生. 假定 $W_{ij} = X_{ij} + U_{ij}$，其中 U_{ij} 服从多元正态分布，均值为 0，协方差阵为 $0.2^2 I_{10}$，I_{10} 是 10×10 单位阵. 取 $\sigma = 0.2$，模拟了三种不同相关系数 $\rho = 0.1, 0.5$ 和 0.8. 选取权函数 $\omega(\cdot)$ 分别为 $\omega(t) = 1$ 和 $\omega(t) = 1/\hat{\sigma}_\varepsilon^2(t)$，获得的 β 估计分别记为 $\hat{\beta}_I$ 和 $\hat{\beta}_{\hat{\omega}}$. 核函数取 $K(t) = K^*(t) = 0.75(1 - t^2)_+$，带宽的选取采用类似 Fan 和 Li[7] 的方法.

本文模拟考虑了两种方法：SCAD 惩罚方法 (SCAD) 和 L_1 惩罚方法 (Lasso). 模拟了三个指标："C"，"I" 和 "GMSE"，其中 "C" 代表零变量正确设为 0 的平均个数，"I" 代表非零变量错误设为 0 的平均个数，"GMSE" 计算公式为

$$\mathrm{GMSE}(\hat{\beta}) = (\hat{\beta} - \beta)^T \big\{\mathrm{Cov}(W) - \Sigma_{uu}\big\}(\hat{\beta} - \beta).$$

模拟结果见表 1. 从表 1 的模拟结果可以看出：

1) SCAD 和 Lasso 都能减少模型的复杂性，但 SCAD 明显优于 Lasso. SCAD 得到的 GMSE 值比 Lasso 小，同时，从 "C" 和 "I" 这两个指标可以看出，SCAD 方法比 Lasso 方法更能有效选出变量. 随着样本量增加，SCAD 越来越接近 Oracle，这与本文的理论结果一致；

2) 对于相同的观测样本来说，当组内相关系数较小时 $\rho = 0.1$，忽略弱相关按照独立数据所得的估计 $\hat{\beta}_I$ 与考虑组内相关所得估计 $\hat{\beta}_{\hat{\omega}}$ 差异不大；当组内相关系数较大时 $\rho = 0.8$，$\hat{\beta}_I$ 的估计效果比 $\hat{\beta}_{\hat{\omega}}$ 差. 由于 $\hat{\beta}_I$ 完全忽视了组内相关性，其结果必然造成估计的效的降低.

4 定理的证明

在证明本文的主要结果之前，首先列出文中所需要的一些正则化条件：

C1: 带宽满足 $h = cN^{-1/5}$，对某个常数 $c > 0$；

C2: 核 $K(\cdot)$ 是对称的概率密度函数，且在它的支撑集 $[-1, 1]$ 上有界变差；

C3:

$$\sup_i \mathrm{E}(\|U_{ij}\|^4) < \infty, \quad \sup_{x, 0 \le t \le 1} \mathrm{E}\big(\varepsilon_{ij}^4 \,|\, X_{ijr} = x, T_{ij} = t\big) < \infty,$$

$$\sup_{0 \le t \le 1} \mathrm{E}\big(X_{ijr}^2 \,|\, T_{ij} = t\big) < \infty, \quad i = 1, 2, \cdots, n, \quad j = 1, 2, \cdots, m_i, \quad r = 1, 2, \cdots, p,$$

其中 X_{ijr} 是 X_{ij} 的第 r 个分量；

C4: 密度函数 $f(t)$ 在 $(0, 1)$ 上是连续可微的且在 $[0, 1]$ 上一致有界，远离零和无穷；

C5: $g(t)$ 和 $\mu_{1r}(t)$ 在 $(0, 1)$ 上是二次连续可微的，$r = 1, 2, \cdots, p$，其中 $\mu_{1r}(t)$ 是 $\mu_1(t)$ 的第 r 个分量；

C6: Γ 是一个正定矩阵，其中 Γ 在 2.2 节中定义；

C7: 方差函数 $\sigma_\varepsilon^2(t)$ 在 t_0 处连续.

引理 4.1 设条件 C1-C5 成立，则

$$\sup_{a \le t \le b} \mathrm{E}\big[\|\mu_1(T_{ij}) - \hat{\mu}_1(T_{ij})\|^2 \,|\, T_{ij} = t\big] = O(n^{-1}h^{-1} + h^4),$$

$$\sup_{a \le t \le b} \mathrm{E}\big[\|g(T_{ij}) - \hat{g}_*(T_{ij})\|^2 \,|\, T_{ij} = t\big] = O(n^{-1}h^{-1} + h^4),$$

其中 $\hat{g}_*(t) = \hat{\mu}_2(t) - \hat{\mu}_1^T(t)\beta$.

引理 4.1 的证明类似文献 [5] 中引理 1 的证明.

表 1: 不同相关系数下变量选择及参数 β 的估计

ρ	n	Method	$\hat{\beta}_I$			$\hat{\beta}_{\hat{\omega}}$		
			C	I	GMSE	C	I	GMSE
0.1	100	Lasso	5.47	0	0.055	5.38	0	0.053
		SCAD	5.70	0	0.051	5.71	0	0.044
		Oracle	6	0	0.022	6	0	0.021
	200	Lasso	5.90	0	0.046	5.92	0	0.037
		SCAD	5.93	0	0.043	5.93	0	0.031
		Oracle	6	0	0.015	6	0	0.013
	400	Lasso	5.97	0	0.038	6	0	0.039
		SCAD	6	0	0.031	6	0	0.031
		Oracle	6	0	0.012	6	0	0.012
0.5	100	Lasso	5.40	0	0.068	5.50	0	0.056
		SCAD	5.50	0	0.061	5.70	0	0.047
		Oracle	6	0	0.026	6	0	0.023
	200	Lasso	5.83	0	0.054	5.84	0	0.044
		SCAD	5.86	0	0.050	5.91	0	0.035
		Oracle	6	0	0.016	6	0	0.015
	400	Lasso	5.90	0	0.048	5.95	0	0.041
		SCAD	5.95	0	0.042	5.98	0	0.036
		Oracle	6	0	0.013	6	0	0.012
0.8	100	Lasso	5.27	0	0.074	5.45	0	0.056
		SCAD	5.43	0	0.071	5.68	0	0.050
		Oracle	6	0	0.028	6	0	0.025
	200	Lasso	5.34	0	0.064	5.81	0	0.046
		SCAD	5.57	0	0.061	5.89	0	0.038
		Oracle	6	0	0.017	6	0	0.016
	400	Lasso	5.69	0	0.058	5.93	0	0.042
		SCAD	5.87	0	0.052	5.96	0	0.039
		Oracle	6	0	0.015	6	0	0.013

定理 2.1 的证明 令 $\alpha_n = n^{-1/2} + a_n$，$\beta = \beta_0 + \alpha_n u$ 且 $\|u\| = C$，其中 C 是充分大的常数. 为了证明定理 2.1，仅需证：对任意给定 ϵ，存在一个大的常数 C，使得对充分大的 n，我们有

$$P\Big\{\min_{\|\beta_0-\beta\|=C\alpha_n}(\beta_0-\beta)^T\Gamma^T\hat{U}^p(\beta)>0\Big\}>1-\epsilon.$$

简单的计算可得

$$n^{-1/2}\hat{U}^p(\beta)=n^{-1/2}\hat{U}(\beta_0)+n^{1/2}\hat{A}(\beta_0-\beta)-n^{1/2}b_\lambda(\beta),$$

令 $\tilde{\mu}_1(t)=\mu_1(t)-\hat{\mu}_1(t)$, $\tilde{g}_*(t)=g(t)-\hat{g}_*(t)$，则

$$\begin{aligned}&\frac{1}{\sqrt{n}}\hat{U}(\beta_0)-\frac{1}{\sqrt{n}}U(\beta_0)\\=&\ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\omega(T_{ij})\tilde{\mu}_1(T_{ij})(\varepsilon_{ij}-U_{ij}^T\beta)+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\omega(T_{ij})\check{W}_{ij}\tilde{g}_*(T_{ij})\\&+\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\sum_{j=1}^{m_i}\omega(T_{ij})\tilde{\mu}_1(T_{ij})\tilde{g}_*(T_{ij})\equiv I_1+I_2+I_3.\end{aligned}$$

结合引理4.1，类似文献 [5] 中引理2的证明可得 $I_v=o_p(1)$, $v=1,2,3$．因此

$$\frac{1}{\sqrt{n}}\hat{U}(\beta_0)=\frac{1}{\sqrt{n}}U(\beta_0)+o_p(1).\tag{7}$$

由引理4.1，可以证明

$$\hat{A}\xrightarrow{p}\Gamma.\tag{8}$$

由 (7) 和 (8) 可得

$$n^{-1/2}\hat{U}^p(\beta)=n^{-1/2}U(\beta_0)+n^{1/2}\Gamma(\beta_0-\beta)-n^{1/2}\big\{b_\lambda(\beta)-b_\lambda(\beta_0)\big\}-n^{1/2}b_\lambda(\beta_0)+o_p(1).$$

如果 $\beta_{0j}\neq 0$，有 $\mathrm{sgn}(\beta_j)=\mathrm{sgn}(\beta_{0j})$．因此

$$q_\lambda(|\beta_j|)\mathrm{sgn}(\beta_j)-q_\lambda(|\beta_{0j}|)\mathrm{sgn}(\beta_{0j})=\big\{q_\lambda(|\beta_j|)-q_\lambda(|\beta_{0j}|)\big\}\mathrm{sgn}(\beta_j).$$

如果 $\beta_{0j}=0$，上式显然成立．那么由中值定理可得

$$\begin{aligned}n^{-1/2}\hat{U}^p(\beta)\ =&\ n^{-1/2}U(\beta_0)+n^{1/2}\Gamma(\beta_0-\beta)\\&+n^{1/2}\mathrm{diag}\big\{q'_\lambda(|\beta_j^*|)\mathrm{sgn}(\beta_j)\big\}(\beta_0-\beta)-n^{1/2}b_\lambda(\beta_0)+o_p(1),\end{aligned}$$

其中 β_j^* 介于 β_j 与 β_{0j} 之间．结合上式，可以证明

$$\begin{aligned}&n^{-1/2}(\beta_0-\beta)^T\Gamma^T\hat{U}^p(\beta)\\=&\ (\beta_0-\beta)^T\Gamma^T\big\{n^{-1/2}\hat{U}(\beta)-n^{1/2}b_\lambda(\beta)\big\}\\=&\ O_p(|\beta_0-\beta|)+n^{\frac{1}{2}}(\beta_0-\beta)^T\Gamma^T\Gamma(\beta_0-\beta)\\&+n^{1/2}(\beta_0-\beta)^T\Gamma^T\mathrm{diag}\big\{q'_\lambda(\beta_j^*)\mathrm{sgn}(\beta_j)\big\}(\beta_0-\beta)-n^{1/2}(\beta_0-\beta)^T\Gamma^Tb_\lambda(\beta)\\=&\ O_p(\alpha_nC)+n^{\frac{1}{2}}(\beta_0-\beta)^T\Gamma^T\Gamma(\beta_0-\beta)+O_p(n^{\frac{1}{2}}\alpha_n^2b_nC^2)+O_p(n^{\frac{1}{2}}\alpha_na_nC).\end{aligned}$$

很容易看出，右边四项中，第一，三，四项由第二项所控制．那么，对任意给定 ϵ，存在一个大的常数 C，使得对充分大的 n，有

$$P\Big\{\min_{\|\beta_0-\beta\|=C\alpha_n}(\beta_0-\beta)^T\Gamma^T\hat{U}^p(\beta)>0\Big\}>1-\epsilon.$$

定理2.2的证明 先证(i). 定理1表明，$\hat{U}^p(\beta)=0$存在$\sqrt{n}$收敛速度的解$\hat{\beta}$. 令$\epsilon_n=Cn^{-1/2}$. 只需证明：对$j=s+1,s+2,\cdots,p$，当$n\to\infty$时，依概率1有

$$\hat{U}^p(\beta_j)>0,\quad 当\quad 0<\beta_j<\epsilon_n, \tag{9}$$

$$\hat{U}^p(\beta_j)<0,\quad 当\quad -\epsilon_n<\beta_j<0. \tag{10}$$

类似于定理2.1的证明可得$\hat{U}(\beta_j)=O_p(\sqrt{n})$. 因此，

$$\hat{U}^p(\beta_j)=n\lambda\{O_p(n^{-1/2}/\lambda)+q_\lambda(|\beta_j|)/\lambda\mathrm{sgn}(\beta_j)\}. \tag{11}$$

由(11)式，结合$n^{1/2}\lambda\to\infty$以及

$$\lim_{n\to+\infty}\inf\lim_{\theta\to0_+}\inf q_\lambda(\theta)/\lambda>0,$$

可知，$\hat{U}^p(\beta_j)$的符号完全由β_j决定. 因此(9),(10)式成立. 即定理2.2的(i)得证.

现考虑$\hat{\beta}^{(1)}$的渐近正态性. 由定理2.1的证明可知

$$n^{-1/2}\hat{U}^p(\hat{\beta})=o_p(1)+n^{-\frac{1}{2}}U(\beta_0)-n^{\frac{1}{2}}\Gamma(\hat{\beta}-\beta_0)-n^{\frac{1}{2}}b_\lambda(\beta_0)-n^{\frac{1}{2}}\Sigma_\lambda(\hat{\beta}-\beta_0). \tag{12}$$

考虑(12)式的前s个方程，有

$$n^{1/2}(\Gamma_{11}+\Sigma_\lambda)\{(\hat{\beta}^{(1)}-\beta_0^{(1)})+(\Gamma_{11}+\Sigma_\lambda)^{-1}b_\lambda(\beta_0)\}$$

$$=n^{-1/2}\begin{pmatrix}U_1(\beta_0)\\ \vdots\\ U_s(\beta_0)\end{pmatrix}+o_p(1)\xrightarrow{\mathcal{L}}N(0,\Xi_{11}).$$

参考文献：

[1] Carroll R J, *et al*. Measurement Error in Nonlinear Models[M]. New York: Chapman & Hall, 2006

[2] 崔恒建. 有重复观测的部分线性EV模型的参数估计[J]. 中国科学(A辑)，2004, 34(4): 467-482
Cui H J. Estimation in partial linear EV models with replicated observations[J]. Science in China (Series A), 2004, 34(4): 467-482

[3] Liang H, *et al*. Estimation in a semiparametric partially linear errors-in-variables model[J]. Annals of Statistics, 1999, 27: 1519-1535

[4] Zhu L X, Cui H J. A semi-parametric regression model with errors in variables[J]. Scandinavian Journal of Statistics, 2003, 30: 429-442

[5] Xue L G, Zhu L X. Empirical likelihood-based inference in a partially linear model for longitudinal data[J]. Science in China Series A: Mathematics, 2008, 51(1): 115-130

[6] 田萍，薛留根. 纵向数据半参数回归模型估计的强相合性[J]. 工程数学学报，2006, 23(2): 369-372
Tian P, Xue L G. Strong consistency of estimators in semiparametric regression model for longitudinal data[J]. Chinese Journal of Engineering Mathematics, 2006, 23(2): 369-372

[7] Fan J Q, Li R Z. New estimation and model selection procedures for semiparametric modeling in longitudinal data analysis[J]. Journal of the American Statistical Association, 2004, 99: 710-723

[8] Liang H, Li R Z. Variable selection for partially linear models with measurement errors[J]. Journal of the American Statistical Association, 2009, 104: 234-248

[9] Fan J Q, Li R Z. Variable selection via nonconcave penalized likelihood and its oracle properties[J]. Journal of the American Statistical Association, 2001, 96: 1348-1360

[10] Li R Z, Liang H. Variable selection in semiparametric regression modeling[J]. Annals of Statistics, 2008, 36: 261-286
[11] Lam C, Fan J Q. Profile-kernel likelihood inference with diverging number of parameters[J]. Annals of Statistics, 2008, 36(5): 2232-2260

Variable Selection in Partially Linear EV Models with Longitudinal Data

YANG Yi-ping[1], XUE Liu-gen[2], CHENG Wei-hu[2]

(1- College of Mathematics and Statistics, Chongqing Technology and Business University, Chongqing 400067; 2- College of Applied Science, Beijing University of Technology, Beijing 100124)

Abstract: The variable selection of the partially linear EV model with longitudinal data is considered in this paper. A variable selection procedure is proposed by the SCAD penalty method. With the appropriate selection of the tuning parameters, the consistency of the variable selection procedure and the Oracle property of the regularized estimators are derived. Numerical simulations are conducted to examine the finite sample performance of the proposed method.

Keywords: partially linear EV model; estimation theory; data analysis; SCAD penalty; Oracle property

Received: 14 Apr 2009. **Accepted:** 24 Feb 2010.

Foundation item: The National Natural Science Foundation of China (10871013); the Ph.D. Program Foundation of the Education Ministry of China (20070005003); the Natural Science Foundation of Beijing Municipality (1102008); PHR (IHLB) and the Research Fund of Chongqing Technology and Business University (20105609).